

Об измерении средней продолжительности безработицы по данным Российского мониторинга экономики и здоровья

В данной статье рассматривается проблема анализа данных Российского мониторинга экономики и здоровья (РМЭЗ) относительно длительности безработицы. Показывается, что результаты предыдущих исследований по данным РМЭЗ подвержены смещению отбора, которое приводит к завышению оценок средней длительности безработицы более чем в 2 раза и к искаженным оценкам функции риска, отражающей, в свою очередь, изменение вероятности выхода из безработицы с течением времени. Предлагается метод анализа, учитывающий смещение отбора. Обновленный подход позволяет получить более реалистичные оценки средней длительности безработицы, которые оказываются согласованными с данными Росстата. Метод может быть использован для оценки эконометрических моделей длительности состояний.

Введение

Одним из наиболее часто используемых индикаторов состояния рынка труда является уровень безработицы. В самом деле доля людей из состава рабочей силы, находящихся в этом неблагоприятном состоянии, представляется вполне разумной мерой измерения тяжести проблемы. Однако одной этой меры явно недостаточно для более-менее полного описания ситуации с безработицей. Корак и Хейц [Corak, Heisz (1995)] приводят следующий пример: уровень безработицы в 10% может наблюдаться в том случае, если в каждом месяце 10% состава рабочей силы теряет работу и находит ее в течение месяца. Но такой же уровень может наблюдаться и тогда, когда эти 10% остаются безработными в течение года, а нового притока в состав безработных нет. Ясно, что во втором случае описывается более проблемная ситуация: попадание в категорию безработных означает для индивида существенное снижение благосостояния.

Чтобы получить более полную картину, используем другой индикатор — среднюю продолжительность безработицы. К сожалению, официальной статистики недостаточно для определения этого показателя. Министерство труда и социального развития РФ предоставляет данные о продолжительности завершенной безработицы среди людей, зарегистрированных органами Государственной службы занятости (ГСЗ). Однако доля безработных, обращающихся в ГСЗ, невелика. Кроме того, выход из числа зарегистрированных в ГСЗ безработных не означает прекращения состояния безработицы: индивид, не нашедший работу с помощью службы занятости, может продолжать поиск ее своими силами. Таким образом, данные Минтруда могут занижать продолжительность безработицы и завышать долю безработных, вышедших из состава рабочей силы.

Другим источником сведений о длительности поиска работы является *обследование населения по проблемам занятости* (ОНПЗ), проводимое Федеральной службой статистики. К сожалению, публикуемые данные ОНПЗ содержат информацию только о продолжительности незавершенного поиска работы (время, в течение которого индивид пробыл безработным на момент опроса). Существуют способы, позволяющие на основе таких данных получить оценку средней продолжительности безработицы (см., например, [Baker, Trivedy (1985)], [Corak, Heisz (1995)]). В частности, Салан [Salant (1997)] показал, что если продолжительность безработицы подчиняется показательному закону, то средняя длительность завершенных периодов безработицы совпадает со средней длительностью незавершенных.

Однако по данным о незавершенных состояниях, нельзя получить ответы на вопросы: чем заканчивается поиск работы? Какая доля безработных находит работу, а какая выходит из состава рабочей силы? Результаты исследований, проведенных различными авторами по РМЭЗ¹, весьма сильно расходятся в оценках средней продолжительности безработицы, но все они значительно превышают оценки Минтруда и ОНПЗ.

В чем же причина этой разницы в оценках? Рассмотрим подробнее, что представляют собой данные РМЭЗ и в чем заключается способ, которым пользовались исследователи для определения продолжительности безработицы.

РМЭЗ — это панель опросов домохозяйств, которые проводятся 1 раз в год или в 2 года. На основании ответов респондента можно определить, является ли он безработным и когда уволился с последнего места работы (если он не имеет работы на момент опроса). Анализируя ответы того же респондента в последующих опросах, можно примерно определить, когда состояние безработицы закончилось, нашел ли респондент работу или прекратил поиск, выйдя из состава рабочей силы. Весьма подробно расчет длительности состояний безработицы описан в работе М. А. Карцевой ([Карцева (2002)], с. 12–16). Проблема заключается в том, что момент опроса, с большой вероятностью, приходится на продолжительные периоды безработицы, в то время как кратковременные периоды между опросами оказываются незамеченными.

Кроме оценки средней длительности безработицы, смещение отбора искажает оценку другой важной для эконометристов характеристики безработицы — функции риска, которая показывает, как с течением времени, проведенного в состоянии безработицы, меняется вероятность покинуть это состояние. Существуют веские основания полагать, что вероятность нахождения работы снижается с течением времени, — как минимум, такой результат должен наблюдаться, если не учитывать неоднородность индивидов. Исследования, утверждающие, что вероятность выхода из безработицы со временем растет (например, Хекман и Сингер [Heckman, Singer (1984)]), учитывают неоднородность индивидов по наблюдаемым признакам (пол, возраст, образование и др.), кроме того, часто пытаются учитывать ненаблюдаемую неоднородность (*unobserved heterogeneity*). Известно, что неоднородность выборки приводит к отрицательному наклону функции риска, т. е. к снижению вероятности выхода ([Neuman (1995)], с. 41).

В то же время оценки функции риска, приводимые в работах Фолей, Грогана и Ван ден Берга [Foley (1997)], [Grogan, van den Berg (1999)], свидетельствуют о немонотонности риска даже без учета неоднородности: вероятность нахождения работы сначала растет, а затем падает, причем вероятность выхода из безработицы в первые месяцы близка к нулю, что весьма странно.

¹ Это прежде всего работы [Foley (1997)], [Grogan, van den Berg (1999)] и [Карцева (2002)].

Опишем статистический эксперимент, результаты которого показывают, что подобные оценки также могут являться следствием смещения отбора. Затем перейдем к изложению метода, позволяющего оценивать закон распределения длительностей по данным РМЭЗ с учетом этого смещения.

1. Основные понятия и методы анализа длительностей

С вероятностной точки зрения индивид может находиться в следующих состояниях:

- 1) работать или находиться в состоянии «занятость»;
- 2) быть безработным, т. е. находиться в состоянии «быть безработным»;
- 3) выйти из состава рабочей силы, т. е. быть в состоянии «не быть в составе рабочей силы» = «не быть рабочей силой».

Длительность некоторого состояния моделируется как неотрицательная случайная величина. Ее, как и другие случайные величины, можно охарактеризовать функцией распределения или функцией плотности (либо вероятностями реализаций, если случайная величина дискретна). Однако в анализе длительностей состояний большее внимание уделяется особым характеристикам, которые будут рассмотрены ниже.

Функция дожития $S(t)$ (часто также называемая функцией надежности) сопоставляет некоторому числу t вероятность того, что случайная величина T примет значение, не меньшее t . Иначе говоря, это вероятность того, что некоторое состояние продлится дольше, чем t единиц времени:

$$S(t) = P(T \geq t) = 1 - F(t), \quad (1)$$

где $F(t)$ — функция распределения случайной величины t .

В качестве характеристики средней продолжительности состояния используется, как правило, математическое ожидание длительности, которое может быть выражено следующим образом:

$$E(T) = \int_0^{\infty} S(t) dt, \quad (2)$$

если T непрерывна.

Для дискретной (точнее, для целочисленной неотрицательной) величины интеграл заменяется на сумму:

$$E(T) = \sum_{t=1}^{\infty} S(t). \quad (3)$$

Кроме математического ожидания длительности, будем также использовать медианную длительность — такую величину T_{med} , что $S(T_{med}) = 0,5$. В своем роде T_{med} также характеризует среднюю продолжительность состояния: половина состояний длится менее T_{med} , и половина оказываются более продолжительными.

Важной характеристикой распределения длительностей является *функция риска* (функция опасности отказов). Для непрерывной случайной величины она определяется следующим образом:

$$\lambda(t) = \frac{\lim_{h \rightarrow 0} P(t \leq T < T + h | T \geq t)}{h}. \quad (4)$$

Можно сказать, что функция риска выражает интенсивность, с которой состояние, длящееся в течение времени t , стремится к прекращению, или вероятность, с которой это состояние прекратится в ближайшее время.

Для дискретной длительности T функция риска определяется как вероятность завершения наблюдаемого состояния в момент t при условии, что это состояние уже «дожило» до этого момента:

$$\lambda(t) = P(T = t | T \geq t). \quad (5)$$

Часто при изучении длительностей представляет интерес именно форма функции риска, так как она отражает характер *временной зависимости* (*duration dependence*).

Говорят, что в момент времени t^* имеет место *положительная* временная зависимость, если

$$\frac{d\lambda(t)}{dt} > 0 \text{ при } t = t^*,$$

т. е. при небольших отклонениях длительностей от t^* для состояния, длящегося в течение большего периода времени, выше вероятность прекращения в ближайшие сроки.

Соответственно существует и *отрицательная* временная зависимость. Если функция риска постоянна, говорят, что временная зависимость отсутствует. Среди непрерывных распределений свойством отсутствия временной зависимости обладает показательное распределение, среди дискретных — геометрическое.

Восстановление закона распределения по имеющимся данным обычно бывает затруднено вследствие цензурирования. Исследователю неизвестна полная длительность некоторых состояний, так как наблюдение за ними было прекращено в некоторый момент, называемый моментом цензурирования (точнее, такой тип цензурирования называется цензурированием справа, *right-censoring*). Например, если на момент последнего наблюдения за индивидом он был безработным, то неизвестно, когда состояние безработицы закончилось. Про цензурированные справа состояния исследователю известно только то, что они продолжались от момента начала до момента цензурирования, т. е. известна их минимальная возможная длительность. Обычно считается, что моменты цензурирования либо детерминированы, либо независимы от длительности состояний. В таких условиях для оценивания функции дожития применим **метод Каплана—Мейера**.

Предположим, моменты прекращения, соответствующие нецензурированным наблюдениям, расположены в порядке возрастания:

$$t_1 < t_2 < \dots < t_K.$$

Количество нецензурированных моментов прекращения K меньше, чем размер выборки n , так как, во-первых, могут присутствовать цензурированные моменты прекращения, а во-вторых, в выборке могли оказаться моменты прекращения с одинаковыми значениями.

Пусть h_j — количество нереализованных моментов цензурирования, т. е. количество наблюдений, моменты прекращения которых действительно наблюдались, с длительностями t_j , $j = 1, \dots, K$. Очевидно, что при отсутствии в выборке длительностей с совпадающими значениями все h_j равны 1.

Введем также величину m_j , показывающую количество наблюдений, цензурированных в момент времени между t_j и t_{j+1} . В данном случае m_K будет означать количество наблюдений, продолжавшихся дольше t_K (длительность, соответствующая наиболее продолжительному нецензурированному наблюдению). Таким образом, величина n_j , которая показывает количество либо незаконченных, либо цензурированных к моменту t_j случайных величин, являющихся моментами прекращения состояний, равна:

$$n_j = \sum_{i \geq j}^K (m_i + h_i). \quad (6)$$

Естественно оценивать функцию риска при значении аргумента t_j как отношение числа нецензурированных наблюдений длительностью t_j к числу наблюдений с длительностями не меньшими t_j :

$$\hat{\lambda}(t_j) = \frac{h_j}{n_j}. \quad (7)$$

Соответственно оценка функции дожития $S(t_j)$ в данном случае будет представлять собой совместную вероятность того, что к моменту t_j событие не завершится:

$$\hat{S}(t_j) = \prod_{i=1}^j \frac{(n_i - h_i)}{n_i} = \prod_{i=1}^j (1 - \hat{\lambda}(t_i)). \quad (8)$$

Это и есть оценка функции дожития Каплана—Мейера. Она не дает информации о поведении функции дожития после последнего наблюдаемого момента прекращения, что не позволяет на ее основе рассчитывать оценку для математического ожидания. Однако, если цензурированных наблюдений не слишком много, метод Каплана—Мейера позволяет оценить медианную длительность.

Если исследователя интересует риск выхода в различные состояния (например, требуется отдельно оценить вероятность трудоустройства, т. е. вероятность перехода в состояние «занятость» и вероятность выхода из состояния «быть рабочей силой», т. е. вероятность перехода в состояние «не быть рабочей силой»), то при оценке риска выхода в некоторое состояние моменты прекращения, в которые происходил выход в другие состояния, считаются моментами цензурирования. Например, если состояние безработицы продолжалось 1 месяц и закончилось выходом индивида из состава рабочей силы, т. е. переходом в состояние «не быть рабочей силой», то при анализе риска выхода в состояние «занятость» это наблюдение будет считаться цензурированным справа в момент «1 месяц».

2. Статистический эксперимент

Чтобы понять, к каким последствиям приводит смещение отбора при анализе длительностей по данным РМЭЗ, был проведен статистический эксперимент, в котором моделировался как истинный процесс перехода между различными состояниями: «занятость», «быть безработным», «не быть в составе рабочей силы» = «не быть рабочей силой», так и отбор наблюдаемых состояний, происходящий в условиях непостоянного наблюдения за индивидами. Опишем нашу модель подробнее.

В момент времени $t = 0$ любой индивид находится в одном из 3 возможных состояний с некоторыми заданными вероятностями. Также предполагается, что каждое из этих состояний началось в тот же момент (это, конечно, неправдоподобная предпосылка, однако, если задать моменты наблюдения достаточно далеко от точки $t = 0$, начальное состояние перестает быть важным).

Длительности каждого состояния генерировались в соответствии с моделью соревнующихся рисков (*competing risks* — [Neumann (1995)], с. 62) следующим образом: существуют две независимые случайные величины, описывающие длительность текущего состояния до момента перехода в одно из двух других состояний (так как для состояния безработицы существуют длительности до выхода «в занятость» и до выхода в неактивность — состояние «не быть рабочей силой»). Обозначим эти величины соответственно T_E и T_N , их функции дожития — $S_E(t)$ и $S_N(t)$, а функции риска — $\lambda_E(t)$ и $\lambda_N(t)$. Наблюдаемая длительность равна минимуму этих двух величин: $T = \min(T_E, T_N)$. Ее функция дожития $S(t)$ равна произведению функций дожития для выхода в каждое из возможных состояний, в нашем случае $S(t) = S_E(t) \cdot S_N(t)$.

Случайные величины для длительностей всех состояний моделировались нами как подчиняющиеся закону распределения Вейбулла.

Устанавливались 3 момента наблюдения (т. е. моделируется ситуация, в которой анализируются данные 3 раундов РМЭЗ). В выборку попадали только те состояния безработицы, которые начались после первого момента наблюдения и которые продолжались на момент второго или третьего моментов наблюдения (раундов опроса). Так поступают Гроган и ван дер Берг, а также М. А. Карцева [Grogan, van den Berg (1999)], [Карцева (2002)] — если включать в выборку состояния, начавшиеся ранее, смещение отбора становится еще сильнее. Если наблюдаемое состояние закончилось до третьего момента, считалось, что моменты его начала и завершения известны с точностью до целого числа, и известно, в какое состояние был осуществлен переход. Если состояние безработицы еще продолжалось к третьему моменту, то наблюдение за ним считалось цензурированным на момент третьего раунда — соответственно состояние, в которое был осуществлен переход, оставалось неизвестным.

Получив таким образом выборку, состоящую частично из нецензурированных, а частично из цензурированных справа наблюдений, мы оценивали распределение длительностей безработицы с помощью метода Каплана—Мейера. Некоторые результаты приведены на графиках (на рис. 1 и 2 представлены графики перечисленных ниже функций).

Из приведенных оценок функции дожития видно, что оценка медианной длительности безработицы сильно завышается. Графики функций риска показывают, что оценки риска являются заниженными, а в области малых длительностей оценки свидетельствуют о положительной временной зависимости, даже когда в реальности временная зависимость отсутствует или является отрицательной. Видимо, именно этот эффект и привел к выявлению немонотонного характера функций риска в работах [Foley (1997)] и [Grogan, van den Berg (1999)].

3. Метод расчета средней продолжительности безработицы по данным РМЭЗ (непрерывная модель)

Рассмотрим случай, когда для анализа используются данные двух раундов опроса. Каждый индивид на момент первого раунда либо не является безработным и не включается в анали-

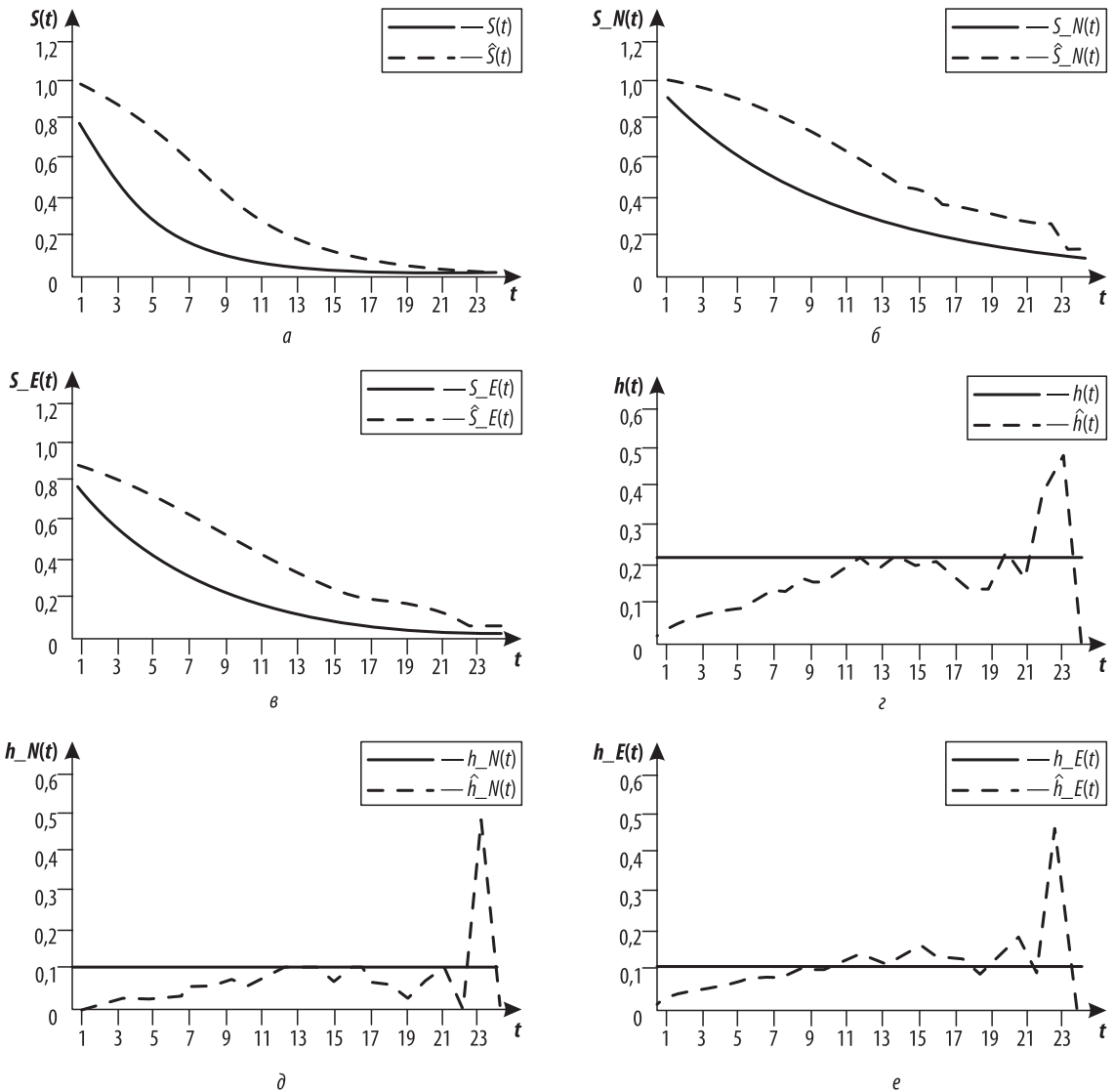


Рис. 1. Отсутствие временной зависимости:

a — истинная функция дожития до выхода из безработицы $S(t)$ и ее оценка $\hat{S}(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); *б* — истинная функция дожития до выхода из рабочей силы $S_N(t)$ и ее оценка $\hat{S}_N(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); *в* — истинная функция дожития до выхода в занятость $S_E(t)$ и ее оценка $\hat{S}_E(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); *г* — истинная функция риска выхода из безработицы $h(t)$ и ее оценка $\hat{h}(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом); *д* — истинная функция риска выхода из состава рабочей силы $h_N(t)$ и ее оценка $\hat{h}_N(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом); *е* — истинная функция риска выхода в занятость $h_E(t)$ и ее оценка $\hat{h}_E(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом)

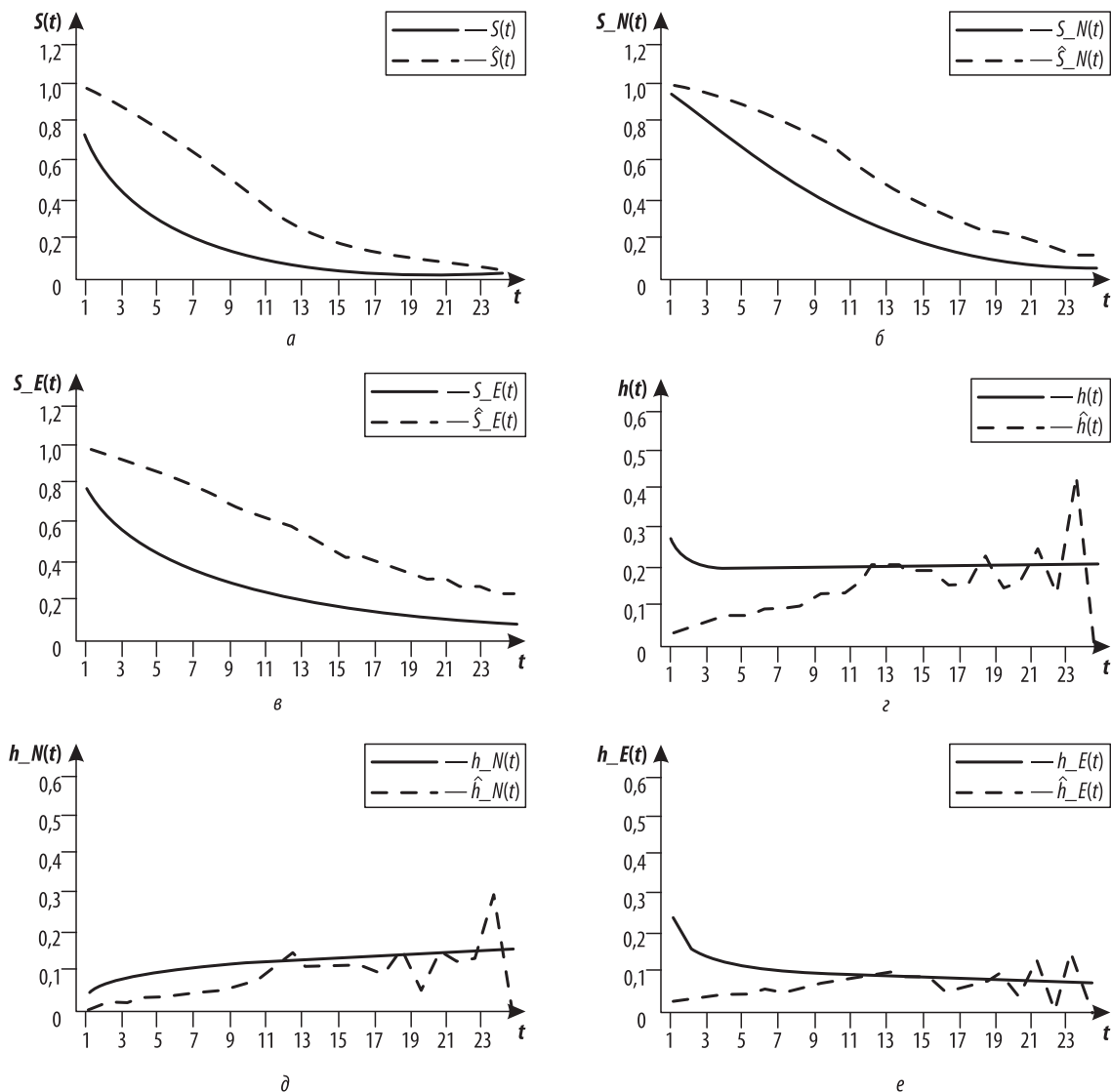


Рис. 2. Положительная временная зависимость для выхода из рабочей силы и отрицательная — для выхода в занятость: а — истинная функция дожития до выхода из безработицы $S(t)$ и ее оценка $\hat{S}(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); б — истинная функция дожития до выхода из рабочей силы $S_N(t)$ и ее оценка $\hat{S}_N(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); в — истинная функция дожития до выхода в занятость $S_E(t)$ и ее оценка $\hat{S}_E(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции дожития, измеряемое числом); г — истинная функция риска выхода из безработицы $h(t)$ и ее оценка $\hat{h}(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом); д — истинная функция риска выхода из состава рабочей силы $h_N(t)$ и ее оценка $\hat{h}_N(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом); е — истинная функция риска выхода в занятость $h_E(t)$ и ее оценка $\hat{h}_E(t)$ (по оси абсцисс откладывается время, измеряемое в условных единицах, по оси ординат — значение функции риска, измеряемое числом)

зируемую выборку, либо находится в состоянии безработицы уже t_i месяцев (i — номер индивида в выборке безработных). На момент второго раунда это состояние либо завершилось, либо продолжается уже $(t_i + l_i)$ месяцев, где l_i — время между опросами i -го индивида в первом и втором раундах. Обратим внимание: если прежнее состояние безработицы завершилось, это вовсе не означает, что на момент второго раунда индивид не является безработным, — он мог покинуть это состояние между опросами и вернуться в него.

Пусть T_i — случайная величина, описывающая продолжительность состояния «быть безработным» у i -го индивида, имеющая функцию дожития $S(t)$. Вероятность того, что i -е состояние на момент второго раунда опроса будет продолжаться при условии, что к моменту первого раунда оно длилось t_i месяцев, равна:

$$P(T_i \geq t_i + l_i | T \geq t_i) = \frac{P(\{T_i \geq t_i + l_i\} \cap \{T_i \geq t_i\})}{P(T_i \geq t_i)} = \frac{P(T_i \geq t_i + l_i)}{P(T_i \geq t_i)} = \frac{S(t_i + l_i)}{S(t_i)}. \quad (9)$$

Соответственно вероятность того, что i -е состояние безработицы закончится к моменту второго раунда опроса, равна:

$$P(T_i < t_i + l_i | T \geq t_i) = 1 - P(T_i \geq t_i + l_i | T \geq t_i) = 1 - \frac{S(t_i + l_i)}{S(t_i)}. \quad (10)$$

Если функция дожития известна с точностью до вектора параметров θ , то эти параметры можно оценить с помощью метода максимального правдоподобия.

В дальнейшем будем использовать следующее обозначение: Y_i — индикатор i -го состояния безработицы на момент второго опроса. $Y_i = 1$, если состояние завершилось, иначе $Y_i = 0$. Тогда задача метода максимального правдоподобия может быть записана следующим образом:

$$L(Y_1, \dots, Y_n; t_1, \dots, t_n; l_1, \dots, l_n; \theta) = \prod_{Y_i=0} \frac{S(t_i + l_i | \theta)}{S(t_i | \theta)} \cdot \prod_{Y_i=1} \left(1 - \frac{S(t_i + l_i | \theta)}{S(t_i | \theta)} \right) \rightarrow \max_{\theta}. \quad (11)$$

Решение задачи даст оценку функции дожития, из которой без труда можно получить и оценку средней продолжительности безработицы.

Рассмотрим более подробно случай, когда длительности подчиняются показательному закону распределения: $S(t) = e^{-\lambda t}$. Тогда средняя продолжительность безработицы выражается через параметр λ следующим образом:

$$E(T) = \frac{1}{\lambda}.$$

Показательный закон обладает свойством отсутствия последствия:

$$P(T < t + x | T \geq t) = P(T < x).$$

Иначе говоря, вероятность того, что некоторое состояние продлится еще несколько месяцев, не зависит от того, сколько оно уже длится (в приложениях, связанных с анализом длительностей состояний, это свойство называется также отсутствием временной зависимости).

Предположение о том, что риск выхода из безработицы постоянен и распределение длительностей подчинено показательному закону, следует из простейшей модели поиска работы (см., например, [Kiefer (1998)]). Несмотря на то что такое ограничение может представлять-

ся жестким, применение показательного распределения является естественным, как минимум, для получения «первого приближения».

При предпосылке об отсутствии временной зависимости вероятности завершения/продолжения i -го состояния на момент второго раунда опроса выражаются следующим образом:

$$P(Y_i = 0) = \frac{S(t_i + l_i)}{S(t_i)} = \frac{e^{-\lambda(t_i + l_i)}}{e^{-\lambda t_i}} = e^{-\lambda l_i}, \quad (12)$$

$$P(Y_i = 1) = 1 - e^{-\lambda l_i}. \quad (13)$$

В этом случае для оценивания параметра λ не нужно обладать информацией о времени, уже проведенном в состоянии безработицы, так как в силу отсутствия последействия это время не влияет на вероятность выхода.

Задачу оценивания λ можно свести к уже известной задаче бинарного выбора — cloglog-регрессии, реализованной в статистических пакетах (например, в Stata). В cloglog-регрессии бинарной величины y на регрессоры x_1, \dots, x_{k-1} вероятности значений y_i выражаются следующим образом:

$$P(y_i = 0) = \exp(-\exp(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_{k-1} x_{k-1,i})); \quad (14)$$

$$P(y_i = 1) = 1 - \exp(-\exp(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_{k-1} x_{k-1,i})). \quad (15)$$

Здесь $\beta_0, \dots, \beta_{k-1}$ — коэффициенты при регрессорах плюс свободный член уравнения. Запишем вероятность (8) в другой форме:

$$P(y_i = 0) = \exp(-\exp(\ln \lambda + \ln l_i)). \quad (16)$$

Очевидно, что выражение (16) сводится к выражению (14) при $k = 2$, $\beta_0 = \ln \lambda$, $\beta_1 = 1$. Таким образом задача оценки параметра сведена к задаче оценки свободного члена в cloglog-регрессии Y на $\ln l$ при том ограничении, что коэффициент при $\ln l$ равен единице. Получив оценку свободного члена $\hat{\beta}_0$, можно рассчитать оценку для параметра λ и среднюю продолжительность безработицы:

$$\hat{E}(T) = \frac{1}{\hat{\lambda}} = e^{-\hat{\beta}_0}. \quad (17)$$

В дальнейшем понадобится также оценка для медианной продолжительности, которую легко получить из условия $S(T_{med}) = 0,5$:

$$\hat{T}_{med} = \frac{\ln 2}{\hat{\lambda}} = \hat{E}(T) \cdot \ln 2. \quad (18)$$

Видно, что медианная продолжительность пропорциональна средней и всегда ниже ее, если длительность описывается показательным законом.

Таким образом, задача оценивания становится достаточно легкой, если не считать затрат сил и времени на подготовку данных. Для того чтобы оценить коэффициент β_0 в пакете Stata, нужно создать переменные y и $\ln l$, содержащие соответственно данные о завершении/продолжении состояний безработицы и логарифм времени между раундами опроса, после чего

ввести команду `cloglog y, offset(ln I)`. Затем следует рассчитать оценку для средней продолжительности безработицы по формуле (17). Можно показать, что если в `cloglog`-регрессию включить объясняющие переменные, то задача оценивания будет являться задачей нахождения коэффициентов в простейшей модели длительности — экспоненциальной регрессии.

Для анализа можно использовать и более двух опросов. Предположим, мы располагаем данными трех раундов. Время между первым и вторым, между вторым и третьим раундами обозначим I_1 и I_2 соответственно. Состояние безработицы, наблюдавшееся в первом раунде и продолжавшееся к моменту первого раунда в течение времени t , могло закончиться между первым и вторым раундами, между вторым и третьим или продолжаться до третьего раунда. Найдем вероятности каждого из этих исходов.

Вероятность того, что i -е состояние закончится до второго раунда опроса:

$$P(T_i < t_i + I_{1,i} | T_i \geq t_i) = 1 - \frac{S(t_i + I_{1,i})}{S(t_i)}. \quad (19)$$

Вероятность того, что i -е состояние завершится между вторым и третьим раундами:

$$P(t_i + I_{1,i} \leq T_i < t_i + I_{1,i} + I_{2,i} | T_i \geq t_i) = P(T_i \geq t_i + I_{1,i} | T_i \geq t_i) \cdot P(T_i < t_i + I_{1,i} + I_{2,i} | T_i \geq t_i + I_{1,i}). \quad (20)$$

Вероятность того, что i -е состояние продлится до третьего раунда:

$$P(T_i \geq t_i + I_{1,i} + I_{2,i} | T_i \geq t_i) = P(T_i \geq t_i + I_{1,i} | T_i \geq t_i) \cdot P(T_i \geq t_i + I_{1,i} + I_{2,i} | T_i \geq t_i + I_{1,i}). \quad (21)$$

Подобным же образом можно расписать вероятности для произвольного количества раундов опроса. Существенным здесь является то, что правую часть выражения (20) можно трактовать и как вероятность совместного наступления двух событий: во-первых, некоторое состояние, продолжавшееся к моменту первого раунда опроса в течение t_i месяцев, будет продолжаться в течение $I_{1,i}$ месяцев до второго раунда; во-вторых, некоторое состояние, продолжавшееся к моменту первого раунда в течение $(t_i + I_{1,i})$ месяцев, закончится в течение $I_{2,i}$ месяцев до второго раунда.

Аналогично выражение (21) можно трактовать как вероятность того, что два состояния, первое из которых продолжалось на момент первого раунда опроса t_i месяцев, а второе — $(t_i + I_{1,i})$ месяцев, будут продолжаться до второго раунда. Как следствие, функция правдоподобия для анализа множества раундов опроса по сути не будет отличаться от функции (11) для двух раундов, хотя количество состояний будет искусственным образом увеличено.

4. Расчет вероятности нахождения работы и выхода из состава рабочей силы в условиях непрерывной модели

Важно знать не только вероятность выхода из безработицы, но и доли безработных, нашедших работу, и безработных прекративших поиск, не найдя ее. Чтобы получить соответствующие оценки, можно использовать модель соревнующихся рисков (*competing risks*).

Пусть существуют две независимые случайные величины: первая — «продолжительность периода безработицы до выхода в состояние «занятость» (T_E) и вторая — «продолжительность периода до выхода из состава рабочей силы — перехода в состояние «не быть рабочей силой» (T_N).

Если $T_E < T_N$, то индивид находит работу, а длительность периода безработицы равна T_E .

Если $T_E > T_N$, то индивид прекращает поиск, проведя в состоянии безработицы время T_N . Соответственно реальная длительность безработицы определяется как:

$$T = \min(T_E, T_N).$$

Это та же самая модель, что использовалась в ранее описанном статистическом эксперименте.

Пусть неизвестен точно момент прекращения состояния безработицы, но известно, закончилось ли это состояние между соседними раундами опроса, и если закончилось, то чем: занятостью или выходом из рабочей силы. Чтобы учесть эту информацию, зададим величину Y_i , как имеющую три возможных значения:

0 — состояние не завершилось;

1 — респондент вышел из состояния, найдя работу;

(-1) — состояние завершилось выходом респондента из состава рабочей силы, т. е. переходом в состояние «не быть рабочей силой».

Теперь нужно выразить вероятности для каждого из значений.

$$P(Y_i = 0) = P(\{T_{E,i} > t_i + l_i\} \cap \{T_{N,i} > t_i + l_i\} | \{T_{E,i} \geq t_i\} \cap \{T_{N,i} \geq t_i\}), \quad (22)$$

$$P(Y_i = 1) = P(\{T_{E,i} < t_i + l_i\} \cap \{T_{E,i} < T_{N,i}\} | \{T_{E,i} \geq t_i\} \cap \{T_{N,i} \geq t_i\}), \quad (23)$$

$$P(Y_i = -1) = P(\{T_{N,i} < t_i + l_i\} \cap \{T_{N,i} < T_{E,i}\} | \{T_{E,i} \geq t_i\} \cap \{T_{N,i} \geq t_i\}). \quad (24)$$

Рассмотрим случай, когда отсутствует временная зависимость для выходов в каждое состояние, т. е. когда обе величины — T_E и T_N — имеют показательное распределение, параметры которого обозначим соответственно λ_E и λ_N . В силу свойства отсутствия последствия, выражения (22)–(24) для этого случая упрощаются:

$$P(Y_i = 0) = P(\{T_{E,i} > l_i\} \cap \{T_{N,i} > l_i\}), \quad (25)$$

$$P(Y_i = 1) = P(\{T_{E,i} < l_i\} \cap \{T_{E,i} < T_{N,i}\}), \quad (26)$$

$$P(Y_i = -1) = P(\{T_{N,i} < l_i\} \cap \{T_{N,i} < T_{E,i}\}). \quad (27)$$

Проще всего получить выражение для $P(Y_i = 0)$, так как события $\{T_{E,i} > l_i\}$ и $\{T_{N,i} > l_i\}$ независимы:

$$P(Y_i = 0) = P(T_{E,i} > l_i) \cdot P(T_{N,i} > l_i) = \exp(-\lambda_E l_i) \exp(-\lambda_N l_i) = \exp(-(\lambda_E + \lambda_N) \cdot l_i). \quad (28)$$

Отметим, что выражение (26) можно записать следующим образом:

$$P(Y_i = 1) = P(\{T_{E,i} < l_i\} \cap \{T_{E,i} < T_{N,i} < l_i\}) + P(\{T_{E,i} < l_i\} \cap \{T_{N,i} > l_i\}) = \\ = P(\{T_{E,i} < l_i\} \cap \{T_{N,i} < l_i\}) \cdot P(\{T_{E,i} < T_{N,i} | T_{E,i} < l_i\} \cap \{T_{N,i} < l_i\}) + P(\{T_{E,i} < l_i\} \cap \{T_{N,i} > l_i\}). \quad (29)$$

Обозначим через $A(\lambda_E, \lambda_N, l_i)$ условную вероятность $P(\{T_{E,i} < T_{N,i} | T_{E,i} < l_i\} \cap \{T_{N,i} < l_i\})$. Можно доказать (см. приложение), что она имеет следующее выражение:

$$A(\lambda_E, \lambda_N, l_i) = \frac{\lambda_E + \lambda_E \exp(-(\lambda_E + \lambda_N) l_i) - (\lambda_N + \lambda_E) \exp(-\lambda_N l_i)}{(\lambda_E + \lambda_N)(1 - \exp(-\lambda_N l_i))(1 - \exp(-\lambda_E l_i))}. \quad (30)$$

Тогда формулу (29) можно записать так:

$$P(Y_i = 1) = P(\{T_{E,i} < l_i\} \cap \{T_{N,i} < l_i\}) \cdot A(\lambda_E, \lambda_N, l_i) + P(\{T_{E,i} < l_i\} \cap \{T_{N,i} > l_i\}) = \\ = (1 - \exp(-\lambda_E l_i))(1 - \exp(-\lambda_N l_i))A(\lambda_E, \lambda_N, l_i) + (1 - \exp(-\lambda_E l_i))\exp(-\lambda_N l_i). \quad (31)$$

Аналогично получаем выражение для $P(Y_i = -1)$:

$$P(Y_i = -1) = (1 - \exp(-\lambda_E l_i))(1 - \exp(-\lambda_N l_i))A(\lambda_N, \lambda_E, l_i) + (1 - \exp(-\lambda_N l_i))\exp(-\lambda_E l_i). \quad (32)$$

Зная выражение вероятностей (25)—(27) через параметры λ_E и λ_N и величину l_i , можно получить оценку параметров методом максимального правдоподобия. Кроме того, обнаружился простой метод, который дает оценки параметров λ_E и λ_N , весьма близкие к оценкам максимального правдоподобия. Алгоритм его применения следующий:

1. Получить оценку параметра λ в модели без учета различных состояний выхода (с помощью cloglog-регрессии).
2. Рассчитать оценки для λ_E и λ_N по следующим формулам:

$$\hat{\lambda}_E = \hat{\lambda} \cdot \frac{\text{Число нашедших работу}}{\text{Число покинувших состояние безработицы}}, \quad (33)$$

$$\hat{\lambda}_N = \hat{\lambda} \cdot \frac{\text{Число вышедших из состава рабочей силы}}{\text{Число покинувших состояние безработицы}}. \quad (34)$$

Различие в оценках, полученных двухшаговым методом, и оценках максимального правдоподобия для проанализированных данных проявлялось не ранее, чем в третьем знаке после запятой. Это практически не влияло на оценку средней продолжительности безработицы и математических ожиданий величин T_E и T_N . Не исключается возможность того, что это расхождение вызвано погрешностью вычислений. Тем не менее свойства оценок двухшагового метода не выяснялись.

5. Дискретная модель длительности

Вышеизложенную непрерывную модель не удалось распространить на тот случай, когда в анализ включаются объясняющие переменные, точнее, не удалось реализовать процедуру оценки такой модели из-за проблем с максимизацией функции правдоподобия. Впрочем, если существует только один вариант выхода, оценивание не составляет труда, однако при анализе безработицы имеет смысл учитывать возможность выхода в занятость и выхода из состава рабочей силы. Несмотря на то что целью данной работы не является изучение влияния индивидуальных характеристик на вероятность выхода из безработицы, понятно, что анализ распределения длительностей — как правило, первый шаг при построении эконометрической модели, включающей экзогенные переменные. Ниже предлагается альтернативный метод оценивания, который позволяет с легкостью включить в модель объясняющие переменные и учесть временную зависимость вероятности выхода из безработицы.

Дискретный подход к моделированию длительности безработицы был, насколько известно автору, впервые использован в статье Никеля [Nickell (1979)] — одной из самых первых работ по данной тематике. Для анализа российских данных дискретный подход был использо-

ван в работе [Foley (1997)]. Применение непрерывных и дискретных моделей часто приводит к одинаковым результатам, так что решение об использовании того или иного подхода, как правило, основывается на удобстве для исследователя. Часто непрерывные модели применяются к данным, которые на самом деле носят дискретный характер (длительности известны, например, с точностью до месяца).

Метод, использованный Никелем, заключается в том, что для состояния «быть безработным», продолжающегося в течение t единиц времени (Никель использовал недельные данные), вероятность выхода из него в ближайшую единицу времени описывается logit-регрессией, где в качестве регрессоров выступают индивидуальные характеристики безработного и некоторая функция от t , описывающая временную зависимость вероятности выхода (в работе [Nickell (1979)] эта зависимость описывалась квадратичной функцией). Кроме logit- часто используется cloglog-регрессия (в таком случае дискретная модель соответствует непрерывной модели пропорциональных рисков — одному из наиболее распространенных подходов к анализу длительностей). Как правило, результаты оценивания по logit- и cloglog- регрессиям схожи.

Так как нас интересуют различные варианты выхода из состояния «быть безработным» — как в состояние «занятость», так и в состояние «не быть рабочей силой», — logit- и cloglog- модели оказываются неподходящими. Естественным расширением logit-регрессии является множественная logit-регрессия (multinomial logit). В таком случае, если моделируется вероятность выхода из безработицы как функция от времени, то вероятности выхода в состояние «занятость» и в состояние «не быть рабочей силой» (обозначим их $p_E(t)$ и $p_N(t)$ соответственно) в течение ближайшего месяца для состояния, продлившегося t месяцев, равны:

$$p_E(t) = \frac{\exp(\varphi_E(t))}{1 + \exp(\varphi_E(t)) + \exp(\varphi_N(t))}, \quad (35)$$

$$p_N(t) = \frac{\exp(\varphi_N(t))}{1 + \exp(\varphi_E(t)) + \exp(\varphi_N(t))}. \quad (36)$$

Обозначим также вероятность остаться в состоянии «быть безработным» до следующего месяца:

$$p_U(t) = 1 - p_E(t) - p_N(t) = \frac{1}{1 + \exp(\varphi_E(t)) + \exp(\varphi_N(t))}. \quad (37)$$

Здесь $\varphi_E(t)$ и $\varphi_N(t)$ — некоторые функции, известные с точностью до некоторого набора параметров и характеризующие временную зависимость вероятности выхода в соответствующее состояние. Например, в работе [Nickell (1979)] использовалась квадратичная функция для вероятности выхода из состояния «быть безработным» (варианты выхода там не различались). Можно предложить и другие варианты. Так, для анализа длительности безработицы мы использовали следующие спецификации:

- постоянная:

$$\varphi(t) = \text{const},$$

- квадратичная:

$$\varphi(t) = a_0 + a_1 t + a_2 t^2,$$

- кусочно-постоянная:

$$\varphi(t) = \begin{cases} \varphi_1, & t < t_1 \\ \varphi_2, & t_1 \leq t \leq t_2, \\ \dots & \\ \varphi_p, & t \geq t_p \end{cases}$$

- кусочно-линейная:

$$\varphi(t) = \begin{cases} \frac{t}{t_1}(\varphi_1 - \varphi_0), & t < t_1 \\ \varphi_1 + \frac{(t - t_1)}{(t_2 - t_1)}(\varphi_2 - \varphi_1), & t_1 \leq t < t_2 \\ \dots & \\ \varphi_{p-1} + \frac{(t - t_{p-1})}{(t_p - t_{p-1})}(\varphi_p - \varphi_{p-1}), & t_{p-1} \leq t < t_p \\ \varphi_p, & t \geq t_p; \end{cases}$$

- экспоненциальная:

$$\varphi(t) = a + be^{-\alpha t}.$$

Оценивание постоянной функции проводилось прежде всего для сравнения результатов, полученных по дискретной и непрерывной моделям, так как в непрерывной спецификации оценивались только постоянные функции риска. Квадратичная, кусочно-постоянная и кусочно-линейная функции являются довольно гибкими и позволяют оценить немонокотонную временную зависимость. Экспоненциальная функция описывает монокотонную временную зависимость, при которой вероятность нахождения работы или выхода из состава рабочей силы стремится к некоторому асимптотическому значению.

Функции риска и дожития выглядят следующим образом:

$$\lambda(t) = 1 - p_U(t) = p_E(t) + p_N(t), \quad (38)$$

$$S(t) = \prod_{k=0}^{t-1} p_U(k). \quad (39)$$

Напомним: наша задача состоит в оценивании закона распределения длительностей на основании имеющихся данных о времени, проведенном i -м индивидом в состоянии «быть безработным» на момент первой волны опроса t_i , времени между первой и второй волнами опроса l_i и индикатором выхода Y_i , равным 0 — для состояний, не завершившихся к моменту второго опроса, 1 — для состояний, завершившихся выходом в занятость, -1 — для состояний, завершившихся выходом из рабочей силы, т. е. переходом в состояние «не быть рабочей силой». В нашей модели это задача оценивания параметров функций $\varphi_E(t)$ и $\varphi_N(t)$, для решения которой будем опираться на метод максимального правдоподобия.

Вероятность того, что за время между волнами опросов состояние безработицы не закончится, равна:

$$P(Y_i = 0 | t_i, l_i) = P(T_i > t_i + l_i | T_i \geq t_i) = \frac{S(t_i + l_i)}{S(t_i)} = \prod_{k=0}^{l_i-1} p_U(t_i + k). \quad (40)$$

Вероятность выхода в состояние «занятость» за период времени между волнами опроса складывается из вероятности найти работу в течение первого месяца, вероятности остаться в состоянии «быть безработным» в первом месяце, но найти работу во втором и т.д.:

$$P(Y_i = 1 | t_i, l_i) = \sum_{k=0}^{l_i-1} \left(p_E(t_i + k) \cdot \prod_{j=0}^{k-1} p_U(t_i + j) \right). \quad (41)$$

Аналогично рассчитывается вероятность выхода из состава рабочей силы:

$$P(Y_i = -1 | t_i, l_i) = \sum_{k=0}^{l_i-1} \left(p_N(t_i + k) \cdot \prod_{j=0}^{k-1} p_U(t_i + j) \right). \quad (42)$$

Задача максимизации функции правдоподобия имеет следующий вид:

$$L = \prod_{Y_i=0} P(Y_i = 0 | t_i, l_i) \prod_{Y_i=1} P(Y_i = 1 | t_i, l_i) \prod_{Y_i=-1} P(Y_i = -1 | t_i, l_i) \rightarrow \max_{\theta_E, \theta_N}, \quad (43)$$

где θ_E и θ_N — вектора оцениваемых параметров функций $\varphi_E(t)$ и $\varphi_N(t)$.

Для того чтобы можно было сравнить результаты дискретной и непрерывной моделей, определим отдельно распределение длительностей до выхода в состояние «занятость» и до выхода из состава рабочей силы. Будем обозначать через $\lambda_E(t)$ и $\lambda_N(t)$ условные функции риска и дожития при условии, что выхода из рабочей силы не произойдет (именно их оценки приведены на рис. 3). Так как в данном случае в каждом месяце существуют только два варианта исхода (остаться безработным или найти работу), множественная logit-модель превращается в обычную logit-спецификацию:

$$\lambda_E(t) = \frac{\exp(\varphi_E(t))}{1 + \exp(\varphi_E(t))}. \quad (44)$$

Соответственно функция дожития до выхода в занятость выглядит следующим образом:

$$S_E(t) = \prod_{k=0}^{t-1} (1 - \lambda_E(k)). \quad (45)$$

Аналогично задаются функции риска и дожития до выхода из состава рабочей силы — $S_N(t)$ и $\lambda_N(t)$.

Из оцененной функции дожития легко рассчитать среднюю длительность безработицы:

$$E(T) = \sum_{t=1}^{\infty} t \cdot P(T = t) = \sum_{t=1}^{\infty} S(t). \quad (46)$$

Стоит отметить, что в дискретную модель длительности легко включить и объясняющие переменные, эффект которых будет суммироваться с эффектом временной зависимости.

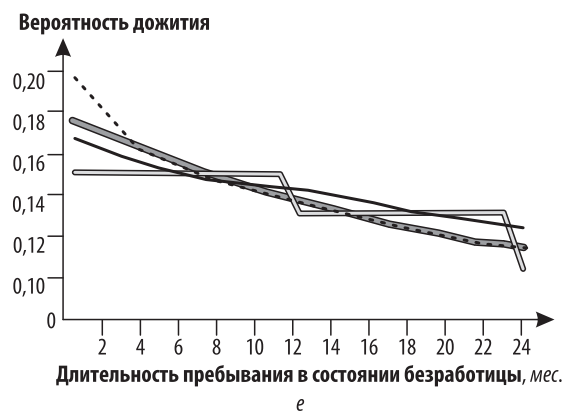
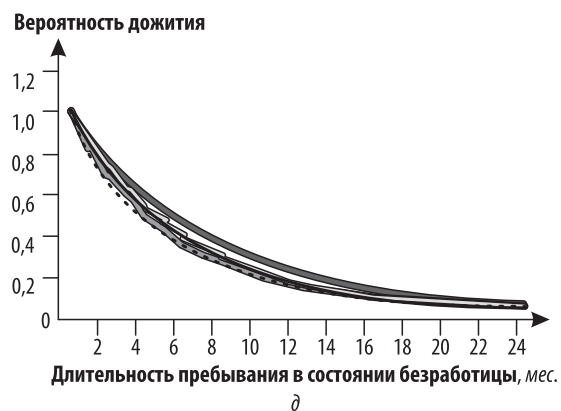
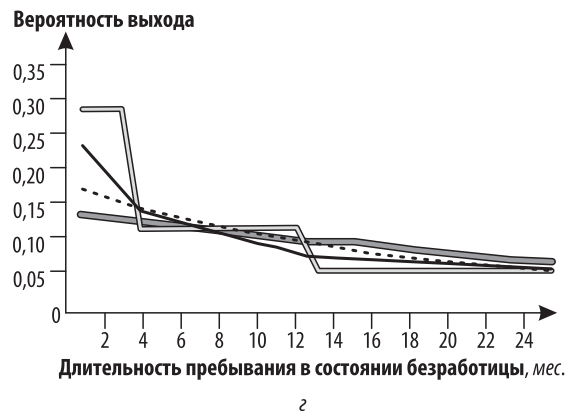
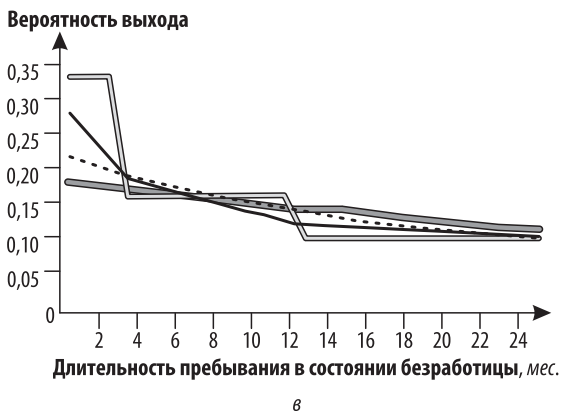
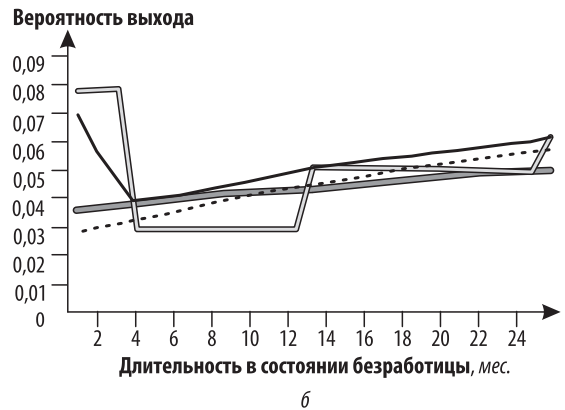
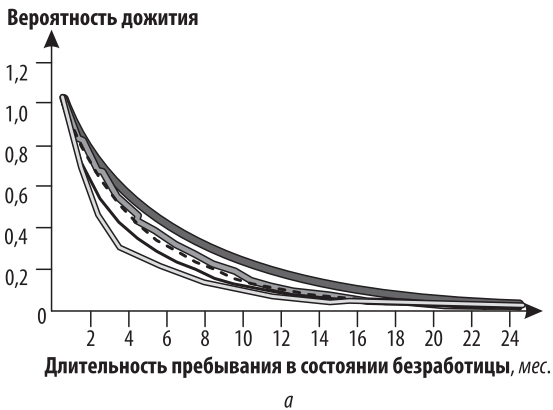
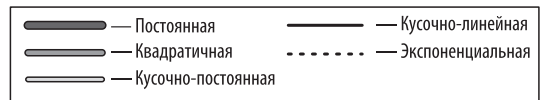


Рис. 3. Оценки функций риска и дожития с учетом временной зависимости:

а — оценки функции дожития, 2000–2005 гг.; б — оценки функции риска выхода из состава рабочей силы, 2000–2005 гг.;
в — оценки функции риска выхода из безработицы, 2000–2005 гг.; г — оценки функции риска выхода в занятость, 2000–2005 гг.;
д — оценки функции дожития, 1994–1996 гг.; е — оценки функции риска выхода из безработицы, 1994–1996 гг.

Например, вероятность нахождения работы в течение месяца можно определить так:

$$p_E(t) = \frac{\exp(\varphi_E(t) + x' \beta_E)}{1 + \exp(\varphi_E(t) + x' \beta_E) + \exp(\varphi_N(t) + x' \beta_N)}, \quad (47)$$

где x' — вектор-строка значений объясняющих переменных (экзогенных факторов, влияющих на вероятность нахождения работы);

вектора β_E и β_N — коэффициенты, определяющие, каким образом значения экзогенных факторов влияют на вероятность нахождения работы и на вероятность выхода из состава рабочей силы.

Аналогично задаются $p_N(t)$ и $p_U(t)$. Отметим, что в вектор объясняющих переменных не стоит включать свободный член, если он присутствует в функциях $\varphi_E(t)$ и $\varphi_N(t)$. Проблем с максимизацией функции правдоподобия не возникает.

6. Данные

В исследовании использовались данные второй волны Российского мониторинга экономики и здоровья (РМЭЗ) за период с 1994 по 2005 г. (V–XIV раунды опроса). Опрос домохозяйств проводился в конце каждого года анализируемого периода, кроме 1997 и 1999 гг. Анализировались состояния безработицы, зафиксированные в раундах V–XIII, данные раунда XIV использовались для того, чтобы выяснить, как завершились (и завершились ли) состояния, наблюдавшиеся в предыдущем раунде.

В анализируемую подвыборку попали только индивиды в возрасте от 18 до 60 лет включительно независимо от пола. Хотя российская служба статистики включает в состав безработных индивидов от 15 до 72 лет, мы намеренно исключили крайние возрастные группы, опасаясь того, что для них закон распределения длительности безработицы может сильно отличаться. По тем же причинам в выборку не включались индивиды, трудоустройству которых мешает плохое состояние здоровья (табл. 1).

Таблица 1

Численность безработных в возрасте 18–60 лет, готовых к работе по состоянию здоровья, в выборке РМЭЗ

Раунд РМЭЗ (год)	Число безработных в выборке, чел.	Число опрошенных, чел.
V (1994)	462*	11 290
VI (1995)	476*	10 666
VII (1996)	525*	10 465
VIII (1998)	617*(593)	10 675
IX (2000)	539*(493)	10 975
X (2001)	514	12 121
XI (2002)	500	12 523
XII (2003)	467	12 656
XIII (2004)	438	12 641

* Выборка построена без учета критерия готовности немедленно приступить к работе.

В определении безработицы использовалась методология Международной организации труда (МОТ). Согласно этой методологии безработным считается индивид:

- не имеющий работы;
- желающий найти и активно ищущий ее;
- готовый приступить к работе в ближайшее время.

Отметим, что некоторые исследователи (Нестерова и Сабирьянова [Нестерова, Сабирьянова (1998)], Гроган и ван ден Берг [Grogan, van den Berg (1999)], Карцева [Карцева (2002)]) включали в категорию безработных также индивидов, которые хотят найти работу, однако не заняты активным поиском. Гроган и ван ден Берг предлагали и другие определения, согласно которым безработными считались индивиды в неоплачиваемом отпуске и даже те, кто испытывает задержки заработной платы. Фолей опирался на *самоопределение* индивидов и включал в анализируемую выборку тех, кто считал себя безработным. В части работ анализировались разные выборки — построенные по классическому определению МОТ и по альтернативным определениям.

Для выявления принадлежности индивида к категории безработных, занятых или не входящих в рабочую силу, использовались ответы на следующие вопросы опросника РМЭЗ:

1. «Ваше основное занятие в настоящее время?»
2. «Хотели бы вы найти работу?»
3. «Вы обращались куда-нибудь или к кому-нибудь в поисках работы в течение последних 30 дней?»
4. «Если бы на прошлой неделе вам предложили подходящую работу, вы смогли бы приступить к ней сразу?»
5. «Какой ответ лучше всего описывает ваше основное занятие в настоящее время?»

Ответ на вопрос 1 использовался для того, чтобы определить, имеет ли респондент работу. Положительные ответы на вопросы 2, 3 и 4 являлись необходимым условием для включения респондента в анализируемую выборку безработных согласно используемому нами определению безработицы. Вопрос 5 был нужен для исключения из анализа людей с плохим здоровьем (среди вариантов ответов предлагался следующий: «не работаю по состоянию здоровья»).

Вопрос 4 был включен в опросники только в 1998 г., поэтому при анализе предыдущих лет не было возможности воспользоваться критерием готовности приступить к работе — индивид классифицировался как безработный, если он удовлетворял всем остальным критериям. Как отмечают К. В. Маркова и С. Ю. Роцин ([Маркова, Роцин (2004)], с. 86), это не должно приводить к значительному искажению результатов, так как более 90% индивидов, отнесенных к безработным по остальным критериям, удовлетворяют и критерию готовности приступить к работе (по данным 1998–2001 гг.). Чтобы иметь возможность сравнить наши результаты с результатами других исследований, для 1998 и 2000 гг. были составлены две выборки безработных — с учетом и без учета критерия готовности к работе.

Чтобы определить, нашел ли респондент работу между двумя раундами опроса, использовались ответы на следующие вопросы:

- «Скажите, пожалуйста, у вас есть сейчас работа, вы находитесь в оплачиваемом или неоплачиваемом отпуске или у вас нет работы?»
- «В каком году и в каком месяце вы ушли с вашей последней работы, перестали работать?»

Если безработный в n -м раунде респондент в следующем раунде опроса указывал, что у него есть работа, считалось, что из состояния «быть безработным» он вышел в состояние

«занятость». Тот же вывод заключался, если респондент указывал, что с последней работы он ушел после раунда n , пусть даже в раунде $(n + 1)$ он опять был безработным (т. е. за промежуток времени между раундами он перешел из безработицы в занятость, а потом обратно). Респондент, не нашедший работу между раундами n и $(n + 1)$, но и не оставшийся в составе безработных, классифицировался как вышедший из рабочей силы. Если в течение периода безработицы индивид стал старше 60 лет, учитывались только те наблюдения за ним, в которых его возраст удовлетворял нашим ограничениям, после 60 лет этот респондент считался выбывшим из рассмотрения (но не вышедшим из состава рабочей силы). Аналогично не рассматривались наблюдения за индивидами младше 18 лет, но как только эти индивиды достигали совершеннолетия, они могли классифицироваться как безработные.

Такой подход не учитывает некоторые варианты переходов. Например, индивид, осуществивший между двумя раундами опроса переход «быть безработным» — «не быть рабочей силой» — «занятость», будет рассматриваться, как вышедший в состояние «занятость», хотя состояние «быть безработным» завершилось выходом в состояние «не быть рабочей силой». Аналогично переход «быть безработным» — «не быть рабочей силой» — «быть безработным» приведет к тому, что выход из состояния «быть безработным» вообще не будет зафиксирован. Такие переходы (как и другие, более сложные, варианты, приводящие к ошибкам классификации) маловероятны в силу того, что состояние «не быть рабочей силой» является наиболее стабильным из рассматриваемых состояний и вероятность выхода из него в течение года довольно мала. Тем не менее был проведен ряд статистических испытаний, которые показали, что смещение оценки в результате ошибок классификации весьма невелико.

При учете временной зависимости считалось, что на вероятность выхода из безработицы влияет время, прошедшее с момента увольнения с последней работы (если месяц увольнения и месяц опроса совпадали, считалось, что респондент пребывает в состоянии безработицы 0 месяцев). Ошибки измерения, связанные с не совпадающими в разных раундах ответами на вопрос о дате увольнения, не учитывались. Они не оказывают существенного влияния на результаты, о чем свидетельствует, в частности, близость оценок средней длительности безработицы в моделях с учетом и без учета временной зависимости.

Время, прошедшее между двумя раундами опроса, рассчитывалось исходя из данных о годе и месяце опроса респондента.

7. Результаты оценивания

Прежде всего, оценивалась непрерывная модель длительности безработицы для двухлетних панелей, составленных из соседних по времени раундов опроса РМЭЗ. Результаты оценивания приведены в табл. 2.

Как видно из табл. 2, наибольшая длительность безработицы наблюдалась в 1996–2000 гг. Частично этот факт можно списать на недостатки используемого метода (как уже говорилось, при увеличении временного промежутка между раундами опроса точность оценок падает из-за ошибок в определении состояния выхода).

Отметим, что это происходит не вследствие недостатка метода самого по себе, а из-за того, что данные РМЭЗ не позволяют точно отслеживать переходы между состояниями «быть безработным» и «не быть рабочей силой». Также отметим, что в этот период времени наблю-

дался и наиболее высокий уровень безработицы (табл. 3), что вполне согласуется с полученными результатами.

Таблица 2

Оценки средней продолжительности безработицы (СПБ) и вероятности выхода из безработицы в течение месяца (ВВБ) в предположении об отсутствии временной зависимости

Годы	Средняя продолжительность безработицы, мес.	Медианная продолжительность безработицы, мес.	Вероятность нахождения работы в течение месяца, %	Вероятность выхода из рабочей силы в течение месяца, %	Вероятность выхода из безработицы в течение месяца, %
1994–1995*	6,3	4,4	10,0	5,2	14,7
1995–1996*	8,6	5,9	7,6	3,7	11,0
1996–1998*	13,6	9,4	4,6	2,6	7,1
1998–2000*	12,3	8,6	5,6	2,3	7,8
2000–2001	7,6	5,3	8,4	4,3	12,3
2001–2002	7,2	5,0	8,3	5,1	13,0
2002–2003	6,2	4,3	9,4	6,0	14,8
2003–2004	6,5	4,5	10,0	4,7	14,3
2004–2005	6,9	4,8	8,1	5,9	13,5

* При анализе первых четырех панелей для определения безработицы не использовался критерий готовности приступить к работе.

Таблица 3

Уровень безработицы и средняя продолжительность поиска работы (незавершенной безработицы), по данным ОНПЗ Госкомстата РФ

Год	Уровень безработицы, %	Среднее время поиска, мес.
1992	5,2	4,4
1993	н/д	5,7
1994	н/д	н/д
1995	9,5	7,4
1996	9,7	8,2
1997	11,8	8,8
1998	13,2	9,1
1999	13,0	9,7
2000	9,8	9,1
2001	8,8	8,2

Год	Уровень безработицы, %	Среднее время поиска, мес.
2002	8,5	8,6
2003	7,8	8,2
2004	7,9	8,5
2005	7,1	8,4
2006	6,7	8,9

Тем не менее наши оценки значительно ниже приводимых в предыдущих исследованиях. Для сравнения перейдем от рассмотрения средней длительности безработицы к медианной длительности. Дело в том, что метод Каплана—Мейера, использованный в предыдущих работах, не позволяет оценить математическое ожидание анализируемой случайной величины, но дает оценку для медианы.

Рассмотрим несколько результатов из имеющихся работ. Приводимая в работе [Foley (1997)]² оценка функции дожития для длительности безработицы в России в 1992–1994 гг. свидетельствует о медианной длительности 14 мес. Гроган и ван ден Берг [Grogan, van den Berg (1999)] дают оценку функции дожития с медианой на уровне 7 мес. (для выборки безработных, производимой по методологии MOT). Согласно Карцевой [Карцева (2002)], медианная длительность безработицы в период 1994–2000 гг. оценивается как 17,9 мес. Как видно, исследователи дают весьма различные оценки.

Оценивание непрерывной модели без учета временной зависимости для выборки 1994–1996 гг. дает оценку медианы в 5,3 мес., а для выборки 1994–2000 гг. — 7,2 мес. К сожалению, мы не смогли оценить длительность безработицы для 1992–1994 гг. из-за отсутствия данных.

Для каждой из двухлетних панелей оценивалась и дискретная модель без временной зависимости (функциями $\varphi_E(t)$ и $\varphi_N(t)$, равными константе). Так как дискретная модель дает оценку среднего числа полных месяцев работы, аналогичный показатель был рассчитан и на основании непрерывной модели, чтобы сделать возможным сравнение двух подходов (табл. 4). Как видно, обе модели дают практически одинаковые результаты.

Оценивание с учетом временной зависимости проводилось по объединенным панелям: отдельно для 1994–1996 гг. и для 2000–2005 гг. Так как для учета временной зависимости необходима информация о дате последнего увольнения, индивиды без опыта работы не входили в эти выборки. Дискретная модель без учета временной зависимости (с функциями $\varphi_E(t)$ и $\varphi_N(t)$, равными константе) оценивалась по объединенным выборкам дважды: до исключения «неопытных» индивидов и после их исключения. Оказалось, что средняя длительность безработицы почти не меняется в зависимости от того, включили «неопытных» индивидов в выборку или нет. Однако оценки, полученные по выборке, состоящей только из «опытных» респондентов, свидетельствуют о том, что для них относительно больше вероятность нахождения работы и меньше вероятность выхода из рабочей силы.

² Приведенные результаты взяты из оценок Каплана—Мейера для функций дожития, приводимых в работах [Foley (1997)] и [Grogan, van den Berg (1999)].

При оценке кусочно заданных функций использовались следующие временные промежутки: 0–11 мес., 12–23 мес., 24–35 мес., 36–47 мес., 48 мес. и более — для выборки 1994–1996 гг., 0–2 мес., 3–11 мес., 12–23 мес., 24–35 мес., 36 мес. и более — для выборки 2000–2005 гг. Результаты, полученные с использованием различных функций временной зависимости $\varphi_E(t)$ и $\varphi_N(t)$, приведены на рис. 3.

Таблица 4

**Среднее число полных месяцев безработицы
согласно непрерывной и дискретной моделям**

Годы	Среднее число полных месяцев безработицы (непрерывная модель)	Среднее число полных месяцев безработицы (дискретная модель)
1994–1995	5,82	6,11
1995–1996	8,08	8,09
1996–1998	13,13	13,07
1998–2000	11,84	11,83
2000–2001	7,10	7,07
2001–2002	6,71	6,67
2002–2003	5,73	5,80
2003–2004	5,99	5,87
2004–2005	6,42	5,92

Оценки закона распределения длительностей при различных спецификациях оказались весьма схожи. Согласно информационному критерию Акаике (табл. 5) наилучшим образом описывает данные модель с экспоненциальными функциями временной зависимости. Оценка математического ожидания неполного числа месяцев поиска работы по этой модели равна 5,5 мес., а оценка медианы — 3,2 мес.³ для выборки 2000–2005 гг. Для ранней выборки эти показатели равны соответственно 6,1 и 3,4 мес.

Как видно, вероятность выхода из безработицы снижается с течением времени — результат типичный, хотя и расходящийся с результатами предыдущих исследований по РМЭЗ. При этом вероятность нахождения работы убывает, а вероятность выхода из состава рабочей силы оценивается различными моделями по-разному. Оценки по кусочно-заданным функциям $\varphi_E(t)$ и $\varphi_N(t)$ свидетельствуют о том, что в первые месяцы эта вероятность высока, затем резко падает, после чего постепенно растет примерно до первоначального уровня (при анализе выборки 2000–2005 гг.). В то же время квадратичная и экспоненциальная формы модели не замечают спада вначале и дают возрастающую функцию риска выхода из рабочей силы. Вероятно, это различие объясняется недостаточной гибкостью квадратичной и экспоненциальной спецификаций.

³ Оценка для медианы получена линейной интерполяцией функции дожития: оцененная функция дожития равна 0,52 для 3-го месяца и 0,42 для 4-го.

Оценки по выборкам 1994–1996 гг. и 2000–2005 гг. оказались весьма схожими, несмотря на то, что меньший объем первой выборки не дает возможности использовать гибкую спецификацию, которая позволила бы точнее оценить временную зависимость в первые месяцы поиска работы. Соответственно все модели свидетельствуют о росте вероятности выхода из состава рабочей силы с течением времени. Модели с кусочно заданной временной зависимостью свидетельствуют о том, что вероятность выхода из безработицы после 2 лет сходится к некоторому устойчивому значению. Экспоненциальная модель дает такие же результаты, что, впрочем, является просто следствием спецификации (так же как квадратичная модель этим результатам противоречит — согласно такой спецификации функция риска обязательно стремится к 1 или 0).

Таблица 5

Качество подгонки для различных спецификаций функций временной зависимости

Спецификация	1994–1996 гг.		2000–2005 гг.	
	Логарифм функции правдоподобия	AIC	Логарифм функции правдоподобия	AIC
Постоянная	–593,37	1190,74	–1570,64	3143,28
Квадратичная	–572,28	1148,57	–1478,41	2962,82
Кусочно-постоянная	–567,70	1139,38	–1470,98	2951,95
Кусочно-линейная	–567,12	1138,25	–1471,77	2953,55
Экспоненциальная	–566,87	1137,74	–1471,95	2949,90

Приводимые оценки средней длительности безработицы не расходятся с данными официальной статистики России. Согласно опубликованным результатам обследования населения по проблемам занятости, средняя продолжительность незавершенных состояний безработицы в период 2000–2005 гг. составляла 8–9 мес. Известно, что в случае отсутствия временной зависимости и неоднородности индивидов средние продолжительности незавершенных и завершенных состояний совпадают [Salant (1977)], если же существует отрицательная временная зависимость, незавершенные состояния в среднем длиннее завершенных ([Carlson, Horrigan (1983)], с. 1144)⁴. Кроме того, в нашу выборку не входили представители крайних возрастных групп и те, кто не работал по состоянию здоровья. Соответственно средняя продолжительность безработицы в нашей выборке должна быть несколько ниже, чем в выборке ОНПЗ, так как для исключенных групп вероятность нахождения работы, скорее всего, относительно низка.

⁴ То, что незавершенные состояния в среднем оказываются такими же или даже более длительными, чем завершенные, может показаться неправдоподобным. Однако стоит учесть, что на продолжительность незавершенных состояний влияет эффект отбора: момент опроса населения, скорее, застанет более продолжительные состояния безработицы.

В то же время полученные оценки несколько выше тех, что приводит Минтруд России. Дело в том, что продолжительность зарегистрированной безработицы ниже продолжительности безработицы, определяемой по методологии МОТ.

Заключение

В настоящей статье показано, что результаты предыдущих исследований безработицы по данным РМЭЗ, относящиеся к средней продолжительности безработицы и характеру временной зависимости, являются следствием смещения отбора, возникающего при попытке применить классические методы анализа длительностей состояний для анализа данных опросов населения. Проведенный статистический эксперимент показал, что игнорирование смещения отбора приводит к значительному завышению оценки средней длительности безработицы и к серьезному искажению функции риска.

В статье предложен метод анализа, который позволяет при работе с данными РМЭЗ получить оценку средней длительности безработицы и функции риска как для выхода в состояние «занятость», так и для перехода в состояние «не быть рабочей силой», а также учесть влияние объясняющих переменных.

Полученные оценки средней продолжительности безработицы согласуются с данными Росстата и Минтруда РФ и значительно ниже оценок, полученных ранее разными исследователями по данным РМЭЗ.

Предложенный метод может быть использован для исследования длительностей других состояний по данным панели опросов населения (анализ брачной стабильности, продолжительности курения и др.).

Список литературы

- Карцева М. А. Влияние образования на вероятность нахождения работы в России (препринт). М., 2002.
- Маркова К. В., Рощин С. Ю. Поиск работы на российском рынке труда. М.: ТЕИС, 2004.
- Нестерова Д., Сабирьянова К. Инвестиции в человеческий капитал в переходный период в России. Серия «Научные доклады». М.: РПЭИ. 1998. № 99. Апрель.
- Савинцева О. С. Анализ длительностей до момента прекращения. Международная научная студенческая конференция. НГУ, 2001.
- Сайт федеральной службы государственной статистики. URL: <http://www.gks.ru>
- Baker G. M., Trivedi P. K. Estimation of Unemployment Duration from Grouped Data: A Comparative Study // *Journal of Labor Economics*. 1985. Vol. 3. № 2. P. 153–174.
- Carlson J. A., Horrigan M. W. Measures of Unemployment Duration as Guides to Research and Policy: Comment // *The American Economic Review*. 1983. Vol. 73. № 5. P. 1143–1150.
- Corak M., Heisz A. Alternative Measures of the Average Duration of Unemployment. Statistics Canada. 1995.
- Foley M. C. Determinants of Unemployment Duration in Russia. Yale University. 1997.
- Grogan L., Van den Berg G. The Duration of Unemployment in Russia. University of Amsterdam, 1999.
- Heckman J., Singer B. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data // *Econometrica*. 1984. Vol. 52. № 2. P. 271–320.
- Kiefer N. M. Economic Duration Data and Hazard Functions // *Journal of Economic Literature*. 1988. Vol. 26. № 2. P. 646–679.

Nickell S. J. Estimating the Probability of Leaving Unemployment // *Econometrica*. 1979. Vol. 47. № 5. P. 1249–1266.

Neumann G. R. Search Models and Duration Data. Handbook of Applied Econometrics. 1995. Ch. 4.

Salant S. W. Search Theory and Duration Data: A Theory of Sorts // *The Quarterly Journal of Economics*. 1977. Vol. 91. № 1. P. 39–57.

ПРИЛОЖЕНИЕ

Доказательство формулы (30)

В силу независимости случайных величин T_E и T_N искомую вероятность можно выразить следующим образом:

$$\begin{aligned} P(T_E < T_N | \{T_E < l\} \cap \{T_N < l\}) &= \\ &= \int_0^l \int_x^l f_E(x | \{T_E < l\}) \cdot f_N(y | \{T_N < l\}) dy dx = \\ &= \int_0^l f_E(x | \{T_E < l\}) \int_x^l f_N(y | \{T_N < l\}) dy dx. \end{aligned}$$

Найдем условную функцию плотности $f_E(x | \{T_E < l\})$ для $x < l$:

$$f_E(x | \{T_E < l\}) = \frac{d}{dx} P(T_E < x | \{T_E < l\}) = \frac{d}{dx} P(T_E < x | \{T_E < l\}) = \frac{d}{dx} \left(\frac{1 - \exp(-\lambda_E x)}{1 - \exp(-\lambda_E l)} \right) = \frac{\lambda_E \exp(-\lambda_E x)}{1 - \exp(-\lambda_E l)}.$$

Аналогично,

$$f_N(y | \{T_N < l\}) = \frac{\lambda_N \exp(-\lambda_N y)}{1 - \exp(-\lambda_N l)}.$$

Таким образом,

$$\begin{aligned} P(T_E < T_N | \{T_E < l\} \cap \{T_N < l\}) &= \int_0^l \frac{\lambda_E \exp(-\lambda_E x)}{1 - \exp(-\lambda_E l)} \int_x^l \frac{\lambda_N \exp(-\lambda_N y)}{1 - \exp(-\lambda_N l)} dy dx = \\ &= \frac{1}{(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))} \int_0^l \lambda_E \exp(-\lambda_E x) \cdot (-\exp(-\lambda_N y)) \Big|_x^l dx = \\ &= \frac{1}{(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))} \int_0^l \lambda_E \exp(-(\lambda_E x)) \cdot (\exp(-\lambda_N x) - \exp(-\lambda_N l)) dx = \\ &= \frac{1}{(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))} \left(\int_0^l \lambda_E \exp(-(\lambda_E + \lambda_N)x) dx - \int_0^l \lambda_E \exp(-\lambda_E x - \lambda_N l) dx \right) = \\ &= \frac{1}{(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))} \left(\left[-\frac{\lambda_E}{\lambda_E + \lambda_N} \exp(-(\lambda_E + \lambda_N)x) \right]_0^l - (-\exp(-(\lambda_E x + \lambda_N l)) \Big|_0^l \right) = \\ &= \frac{1}{(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))} \left(\frac{\lambda_E}{\lambda_E + \lambda_N} (1 - \exp(-(\lambda_E + \lambda_N)l)) + \exp(-(\lambda_E + \lambda_N)l) - \exp(-\lambda_N l) \right) = \\ &= \frac{\lambda_E + \lambda_N \exp(-(\lambda_E + \lambda_N)l) - (\lambda_E + \lambda_N) \exp(-\lambda_N l)}{(\lambda_E + \lambda_N)(1 - \exp(-\lambda_E l))(1 - \exp(-\lambda_N l))}, \end{aligned}$$

что и требовалось доказать.